WHAT IS CLAIMED IS:

1. A method of determining whether a test sample, having test data T, is categorized in one of a number n of classes wherein n is 2 or more, comprising:

extracting a plurality of emerging patterns from a training data set D that has at least one instance of each of said n classes of data;

creating n lists, wherein:

an ith list of said n lists contains a frequency of occurrence, $f_i(m)$, of each emerging pattern $EP_i(m)$ from said plurality of emerging patterns that has a non-zero occurrence in an ith class of data;

using a fixed number, k, of emerging patterns, wherein k is substantially less than a total number of emerging patterns in the plurality of emerging patterns, calculating n scores; wherein:

an ith score of said n scores is derived from the frequencies of k emerging patterns in said ith list that also occur in said test data; and

deducing which of said n classes of data the test data is categorized in, by selecting the highest of said n scores.

2. The method of claim 1, additionally comprising:

if there is more than one class with the highest score, deducing which of said n classes of data the test data is categorized in by selecting the largest of the classes of data having the highest score.

3. The method of claim 1 or 2, wherein:

said k emerging patterns of the ith list that occur in said test data have the highest frequencies of occurrence in said ith list amongst all those emerging patterns of said ith list that occur in said test data, for all i.

4. The method of any one of the preceding claims, wherein:

emerging patterns in the ith list are ordered in descending order of said frequency of occurrence in said ith class of data, for all i.

5. The method of any one of the preceding claims, wherein the ith list has a length l_i, and k is a fixed percentage of the smallest l_i.

6. The method of any one of claims 1 to 4, wherein the ith list has a length l_i, and k is a fixed percentage of $\sum_{i=1}^{n} l_i$

7. The method of any one of claims 1 to 4, wherein the ith list has a length l_i, and k is a fixed percentage of any l_i.

8. The method of any one of claims 5 to 7, wherein said fixed percentage is from about 1% to about 5% and k is rounded to a nearest integer value.

9. The method of any one of the preceding claims, wherein n = 2.

10. The method of any one of claims 1 to 8, wherein n = 3 or more.

11. A method of determining whether a test sample, having test data T, is categorized in a first class or a second class, comprising:

extracting a plurality of emerging patterns from a training data set D that has at least one instance of a first class of data and at least one instance of a second class of data;

creating a first list and a second list wherein:

said first list contains a frequency of occurrence, $f_1(m)$, of each emerging pattern $EP_1(m)$ from said plurality of emerging patterns that has a non-zero occurrence in said first class of data; and

said second list contains a frequency of occurrence, $f_2(m)$, of each emerging pattern $EP_2(m)$ from said plurality of emerging patterns that has a non-zero occurrence in said second class of data;

using a fixed number, k, of emerging patterns, wherein k is substantially less than a total number of emerging patterns in the plurality of emerging patterns, calculating:

a first score derived from the frequencies of k emerging patterns in said first list that also occur in said test data, and

a second score derived from the frequencies of k emerging patterns in said second list that also occur in said test data; and

deducing whether the test data is categorized in the first class of data or in the second class of data by selecting the higher of said first score and said second score.

12. The method of claim 11, additionally comprising:

if said first score and said second score are equal, deducing whether the test sample is categorized in the first class of data or in the second class of data by selecting the larger of the first or the second class of data.

13. The method of claim 11 or 12, wherein:

said k emerging patterns of said first list that occur in said test data have the highest frequencies of occurrence in said first list amongst all those emerging patterns of said first list that occur in said test data; and

said k emerging patterns of said second list that occur in said test data have the highest frequencies of occurrence in said second list amongst all those emerging patterns of said second list that occur in said test data.

14. The method of any one of claims 11 to 13, wherein:

emerging patterns in said first list are ordered in descending order of said frequency of occurrence in said first class of data, and

emerging patterns in said second list are ordered in descending order of said frequency of occurrence in said second class of data.

15. The method of any one of claims 11 to 14, additionally comprising:

creating a third list and a fourth list, wherein:

said third list contains a frequency of occurrence, $f_1(i_m)$, in said first class of data of each emerging pattern i_m from said plurality of emerging patterns that has a non-zero occurrence in said first class of data and which also occurs in said test data; and

said fourth list contains a frequency of occurrence, $f_2(j_m)$, in said second class of data of each emerging pattern j_m from said plurality of emerging patterns that has

a non-zero occurrence in said second class of data and which also occurs in said test data; and wherein

emerging patterns in said third list are ordered in descending order of said frequency of occurrence in said first class of data, and

emerging patterns in said fourth list are ordered in descending order of said frequency of occurrence in said second class of data.

16. The method of claim 15, wherein:

said first score is given by: $\sum_{m=1}^{k} \frac{f_1(i_m)}{f_1(m)}\Big|_{EP_1(i_m)\in T}$; and

said second score is given by: $\sum_{m=1}^{k} \frac{f_2(j_m)}{f_2(m)}\Big|_{EP_2(j_m)\in T}$.

17. The method of any one of claims 11 to 16, wherein said first list has a length l_1, and said second list has a length l_2, and k is a fixed percentage of whichever of l_1 and l_2 is smaller.

18. The method of any one of claims 11 to 16, wherein said first list has a length l_1, and said second list has a length l_2, and k is a fixed percentage of a sum of l_1 and l_2.

19. The method of any one of claims 11 to 16, wherein said first list has a length l_1, and said second list has a length l_2, and k is a fixed percentage of any one of l_1 or l_2.

20. The method of any one of claims 17 to 19, wherein said fixed percentage is from about 1% to about 5% and k is rounded to a nearest integer value.

21. The method of any one of the preceding claims, wherein k is from about 5 to about 50.

22. The method of claim 21, wherein k is about 20.

23. The method of any one of the preceding claims, wherein each emerging pattern is expressed as a conjunction of conditions.

24. The method of any one of the preceding claims, wherein only left boundary emerging patterns are used.

25. The method of any one of claims 1 to 23, wherein only plateau emerging patterns are used.

26. The method of claim 25 wherein only the most specific plateau emerging patterns are used.

27. The method of any one of the preceding claims, wherein each of said emerging patterns has a growth rate larger than a threshold, □.

28. The method of claim 27 wherein said threshold is from about 2 to about 10.

29. The method of any one of the preceding claims, wherein each of said emerging patterns has a growth rate of ∞.

30. The method of any one of the preceding claims, additionally comprising discretizing said data set, before said extracting.

31. The method of claim 30, wherein said discretizing utilizes an entropy-based method.

32. The method of claim 30 or 31, additionally comprising applying a method of correlation based feature selection to said data set, after said discretizing.

33. The method of claim 30, 31 or 32, additionally comprising applying a chi-squared method to said data set, after said discretizing.

34. The method of any one of the preceding claims, wherein said data set comprises gene expression data.

35. The method of claim 34, wherein said gene expression data has been acquired from a micro-array apparatus.

36. The method of any one of the preceding claims, wherein at least one class of data corresponds to data for a first type of cell and at least another class of data corresponds to data for a second type of cell.

37. The method of claim 36, wherein said first type of cell is a normal cell and said second type of cell is a cancerous cell.

38. The method of any one of the preceding claims, wherein at least one class of data corresponds to data for a first population of subjects and at least another class of data corresponds to data for a second population of subjects.

39. The method of any one of claims 1 to 33, wherein said data set comprises patient medical records.

40. The method of any one of claims 1 to 33, wherein said data set comprises financial transactions.

41. The method of any one of claims 1 to 33, wherein said data set comprises census data.

42. The method of any one of claims 1 to 33, wherein said data set comprises characteristics of an item selected from the group consisting of: a foodstuff; an article of manufacture; and a raw material.

43. The method of any one of claims 1 to 33, wherein said data set comprises environmental data.

44. The method of any one of claims 1 to 33, wherein said data set comprises meteorological data.

45. The method of any one of claims 1 to 33, wherein said data set comprises

characteristics of a population of organisms.

46. The method of any one of claims 1 to 33, wherein said data set comprises marketing data.

47. A computer program product for determining whether a test sample, for which there exists test data, is categorized in a first class or a second class, wherein the computer program product is for use in conjunction with a computer system, the computer program product comprising:

a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising:

at least one statistical analysis tool;

at least one sorting tool; and

control instructions for:

accessing a data set that has at least one instance of a first class of data and at least one instance of a second class of data;

extracting a plurality of emerging patterns from said data set;

creating a first list and a second list wherein, for each of said plurality of emerging patterns:

said first list contains a frequency of occurrence, $f_i^{(1)}$, of each emerging pattern i from said plurality of emerging patterns that has a non-zero occurrence in said first class of data, and

said second list contains a frequency of occurrence, $f_i^{(2)}$, of each emerging pattern i from said plurality of emerging patterns that has a non-zero occurrence in said second class of data;

using a fixed number, k, of emerging patterns, wherein k is substantially less than a total number of emerging patterns in the plurality of emerging patterns, calculating:

a first score derived from the frequencies of k emerging patterns in said first list that also occur in said test data, and

a second score derived from the frequencies of k emerging patterns in said second list that also occur in said test data; and

deducing whether the test sample is categorized in the first class of data or in the second class of data by selecting the higher of the first score and the second score.

48. The computer program product of claim 47, additionally comprising instructions for:

if said first score and said second score are equal, deducing whether the test sample is categorized in the first class of data or in the second class of data by selecting the larger of the first or the second class of data.

49. The computer program product of claim 47 or 48, wherein:

said k emerging patterns of said first list that occur in said test data have the highest frequencies of occurrence in said first list amongst all those emerging patterns of said first list that occur in said test data; and

said k emerging patterns of said second list that occur in said test data have the highest frequencies of occurrence in said second list amongst all those emerging patterns of said second list that occur in said test data.

50. The computer program product of any one of claims 47 to 49, further comprising control instructions for:

ordering emerging patterns in said first list in descending order of said frequency of occurrence in said first class of data, and

ordering emerging patterns in said second list in descending order of said frequency of occurrence in said second class of data.

51. The computer program product of any one of claims 47 to 50, additionally comprising instructions for:

creating a third list and a fourth list, wherein:

said third list contains a frequency of occurrence, $f_1(i_m)$, in said first class of data of each emerging pattern i_m from said plurality of emerging patterns that has a non-zero occurrence in said first class of data and which also occurs in said test data; and

said fourth list contains a frequency of occurrence, $f_2(j_m)$, in said second class

of data of each emerging pattern j_m from said plurality of emerging patterns that has a non-zero occurrence in said second class of data and which also occurs in said test data; and wherein

emerging patterns in said third list are ordered in descending order of said frequency of occurrence in said first class of data, and

emerging patterns in said fourth list are ordered in descending order of said frequency of occurrence in said second class of data.

52. The computer program product of claim 51, further comprising instructions for calculating:

said first score according to the formula: $\sum_{m=1}^{k} \frac{f_1(i_m)}{f_1(m)}\Big|_{EP_1(i_m)\in T}$; and

said second score according to the formula: $\sum_{m=1}^{k} \frac{f_2(j_m)}{f_2(m)}\Big|_{EP_2(j_m)\in T}$.

53. The computer program product of any one of claims 47 to 52, wherein k is from about 5 to about 50.

54. The computer program product of any one of claims 47 to 53, wherein only left boundary emerging patterns are used.

55. The computer program product of any one of claims 47 to 54, wherein each of said emerging patterns has a growth rate of ∞.

56. The computer program product of any one of claims 47 to 55, wherein said data set comprises data selected from the group consisting of: gene expression data, patient medical records, financial transactions, census data, characteristics of an article of manufacture, characteristics of a foodstuff, characteristics of a raw material, meteorological data, environmental data, and characteristics of a population of organisms.

57. A system for determining whether a test sample, for which there exists test data, is categorized in a first class or a second class, the system comprising:

at least one memory, at least one processor and at least one user interface, all of which are connected to one another by at least one bus;

wherein said at least one processor is configured to:

access a data set that has at least one instance of a first class of data and at least one instance of a second class of data;

extract a plurality of emerging patterns from said data set;

create a first list and a second list wherein, for each of said plurality of emerging patterns:

said first list contains a frequency of occurrence, $f_i^{(1)}$, of each emerging pattern i from said plurality of emerging patterns that has a non-zero occurrence in said first class of data, and

said second list contains a frequency of occurrence, $f_i^{(2)}$, of each emerging pattern i from said plurality of emerging patterns that has a non-zero occurrence in said second class of data;

use a fixed number, k, of emerging patterns, wherein k is substantially less than a total number of emerging patterns in the plurality of emerging patterns, to calculate:

a first score derived from the frequencies of k emerging patterns in said first list that also occur in said test data, and

a second score derived from the frequencies of k emerging patterns in said second list that also occur in said test data; and

deduce whether the test sample is categorized in the first class of data or in the second class of data by selecting the higher of the first score and the second score.

58. The system of claim 57, wherein said processor is additionally configured to:

if said first score and said second score are equal, deduce whether the test sample is categorized in the first class of data or in the second class of data by selecting the larger of the first or the second class of data.

59. The system of claim 57 or 58, wherein:

said k emerging patterns of said first list that occur in said test data have the highest frequencies of occurrence in said first list amongst all those emerging patterns of said first

list that occur in said test data; and

said k emerging patterns of said second list that occur in said test data have the highest frequencies of occurrence in said second list amongst all those emerging patterns of said second list that occur in said test data.

60. The system of claim 57, 58 or 59, wherein said processor is additionally configured to:

order emerging patterns in said first list in descending order of said frequency of occurrence in said first class of data, and

order emerging patterns in said second list in descending order of said frequency of occurrence in said second class of data.

61. The system of any one of claims 57 to 60, wherein said processor is additionally configured to:

create a third list and a fourth list, wherein:

said third list contains a frequency of occurrence, $f_1(i_m)$, in said first class of data of each emerging pattern i_m from said plurality of emerging patterns that has a non-zero occurrence in said first class of data and which also occurs in said test data; and

said fourth list contains a frequency of occurrence, $f_2(j_m)$, in said second class of data of each emerging pattern j_m from said plurality of emerging patterns that has a non-zero occurrence in said second class of data and which also occurs in said test data; and wherein

emerging patterns in said third list are ordered in descending order of said frequency of occurrence in said first class of data, and

emerging patterns in said fourth list are ordered in descending order of said frequency of occurrence in said second class of data.

62. The system of claim 61, wherein said processor is additionally configured to calculate:

said first score according to the formula: $\sum_{m=1}^{k} \frac{f_1(i_m)}{f_1(m)}\Big|_{EP_1(i_m)\in T}$; and

said second score according to the formula: $\sum_{m=1}^{k} \frac{f_2(j_m)}{f_2(m)}\Big|_{EP_2(j_m)\in T}$.

63. The system of any one of claims 57 to 62, wherein k is from about 5 to about 50.

64. The system of any one of claims 57 to 63, wherein only left boundary emerging patterns are used.

65. The system of any one of claims 57 to 64, wherein each of said emerging patterns has a growth rate of ∞.

66. The system of any one of claims 57 to 65, wherein said data set comprises data selected from the group consisting of: gene expression data, patient medical records, financial transactions, census data, characteristics of an article of manufacture, characteristics of a foodstuff, characteristics of a raw material, meteorological data, environmental data, and characteristics of a population of organisms.

67. A method of determining whether a sample cell is cancerous, comprising:

extracting a plurality of emerging patterns from a data set that comprises gene expression data for a plurality of cancerous cells and a gene expression data for a plurality of normal cells;

creating a first list and a second list wherein:

said first list contains a frequency of occurrence, $f_i^{(1)}$, of each emerging pattern i from said plurality of emerging patterns that has a non-zero occurrence in said cancerous cells, and

said second list contains a frequency of occurrence, $f_i^{(2)}$, of each emerging pattern i from said plurality of emerging patterns that has a non-zero occurrence in said normal cells;

using a fixed number, k, of emerging patterns, wherein k is substantially less than a total number of emerging patterns in the plurality of emerging patterns, calculating:

a first score derived from the frequencies of k emerging patterns in said first list that also occur in said test data, and

a second score derived from the frequencies of k emerging patterns in said second list that also occur in said test data; and

deducing whether the sample cell is cancerous if said first score is higher than said second score.

68. A method of determining whether a test sample, having test data T, is categorized in one of a number of classes, substantially as hereinbefore described with reference to and as illustrated in the accompanying drawings.

69. The computer program product of any one of claims 47 to 56, operable according to the method of any one of claims 1 to 46, 67 and 68.

70. A computer program product operable according to the method of any one of claims 1 to 46, 67 and 68.

71. A computer program product for determining whether a test sample, for which there exists test data, is categorized in one of a number of classes, constructed and arranged to operate substantially as hereinbefore described with reference to and as illustrated in the accompanying drawings.

72. The system of any one of claims 57 to 66, operable according to the method of any one of claims 1 to 46, 67 and 68.

73. A system for determining whether a test sample, for which there exists test data, is categorized in one of a number of classes, constructed and arranged to operate substantially as hereinbefore described with reference to and as illustrated in the accompanying drawings.

74. A system operable according to the method of any one of claims 1 to 46, 67 and 68.

75. The system of any one of claims 57 to 66 and 71 to 73, for use with the computer program product of any one of claims 47 to 56 and 69 to 71.